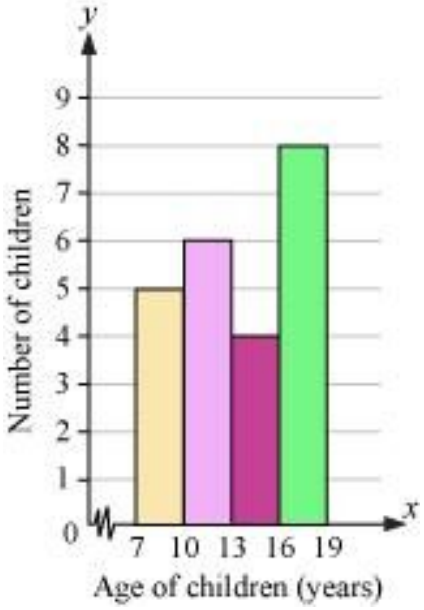
Graphical Representation

- **Histogram**

  A histogram is a bar graph that is used to represent grouped data. In a histogram, the class intervals are represented on the horizontal axis and the heights of the bars represent frequency. Also, there is no gap between the bars in a histogram.

Class interval (Age of children)	**Tally mark**	**Frequency (Number of children)**
7 – 10	𝍸	5
10 – 13	𝍸 |	6
13 – 16	||||	4
16 – 19	𝍸 |||	8

The above frequency distribution table can be displayed in a histogram as follows:



In a histogram, a broken line can be used along the horizontal axis to indicate that the numbers between 0 to7 are not included.

- The observation with maximum frequency is called **mode**. When data is grouped into classes, mode can be obtained by drawing histogram.

Example:

The following table shows the class intervals and the frequency corresponding to them.

Class Interval	**40 − 69**	**70 − 99**	**100 − 129**	**130 − 159**	**160 − 189**
Frequency	**12**	**15**	**24**	**16**	**21**

Find the mode of the given data geometrically.

Solution:

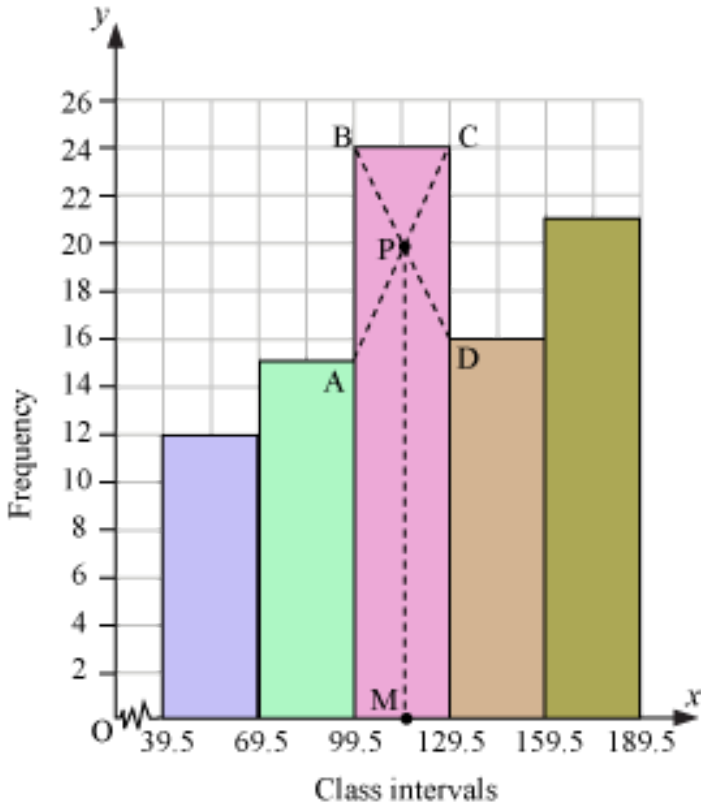
The given frequency distribution is discontinuous. To convert it into continuous distribution, we have to subtract $\frac{1}{2} = 0.5$ from the lower limits and add 0.5 to higher limits of each class interval.

Now, the continuous frequency distribution table of the given data is as follows:

Class Interval	**Frequency**

39.5 − 69.5	12
69.5 − 99.5	15
99.5 − 129.5	24
129.5 − 159.5	16
159.5 − 189.5	21

To find the mode of the above data geometrically, first of all we have to draw its histogram by choosing 1 cm along x-axis = 30 (class-intervals) and 1 cm = 2 (frequencies). In the highest rectangle (class interval 99.5 − 129.5), we will draw two straight lines AC and BD from corners of the rectangles on either side of the highest rectangle to the opposite corners of the highest rectangle. Let P be the intersection of the lines AC and BD. Now, we will draw a vertical line through the point P that cuts the x-axis at M.



The point M represents the value 115 (approximately) on x-axis. Therefore, the mode of the given data is 115 (approximately).

- **Graphical representation of cumulative frequency distribution Ogive**

- **OGIVE (of the less- than type)**

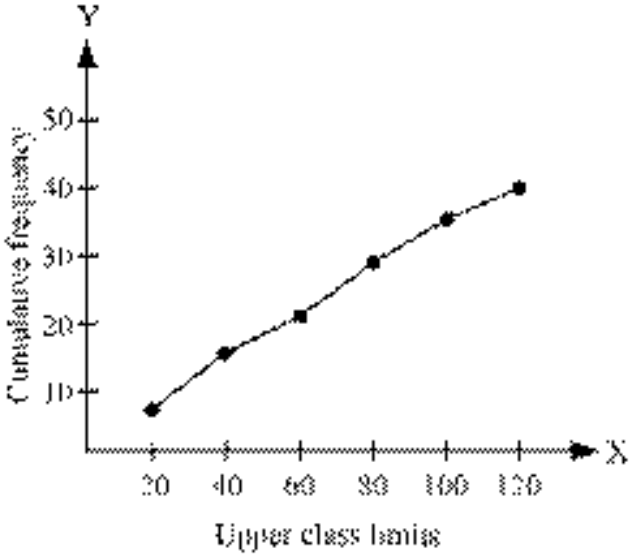
Example 1: Draw ogive of the less-than type for the given distribution.

Class interval	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100	100 – 120
Frequency	7	8	6	8	6	5

Solution: The cumulative frequency distribution for the given data can be found as:

Class interval	**Upper class limit**	**Frequency**	**Cumulative frequency**
0 – 20	20	7	7
20 – 40	40	8	15
40 – 60	60	6	21
60 – 80	80	8	29
80 – 100	100	6	35
100 – 120	120	5	40

By taking the horizontal axis as the upper class limit and the vertical axis as the corresponding cumulative frequency, we can plot the cumulative frequency for each upper class limit.
Then, the required ogive (of the less-than type) is obtained as:



- **OGIVE (of the more-than type)**

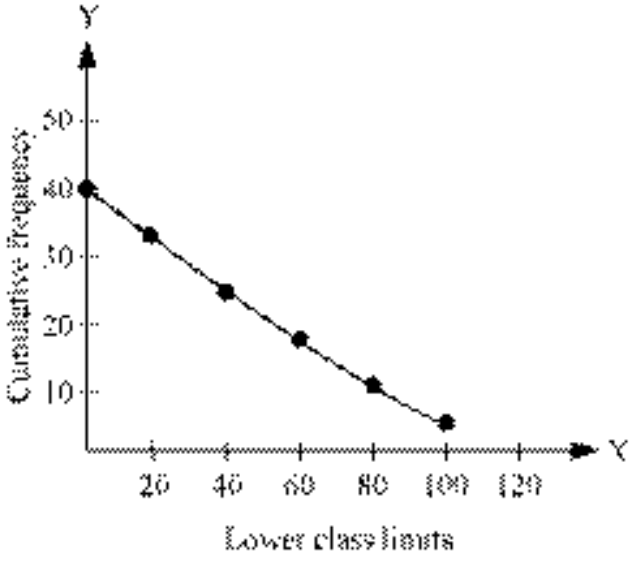
Example 2: Draw ogive of the more-than type for the following distribution.

Class interval	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100	100 – 120
Frequency	7	8	6	8	6	5

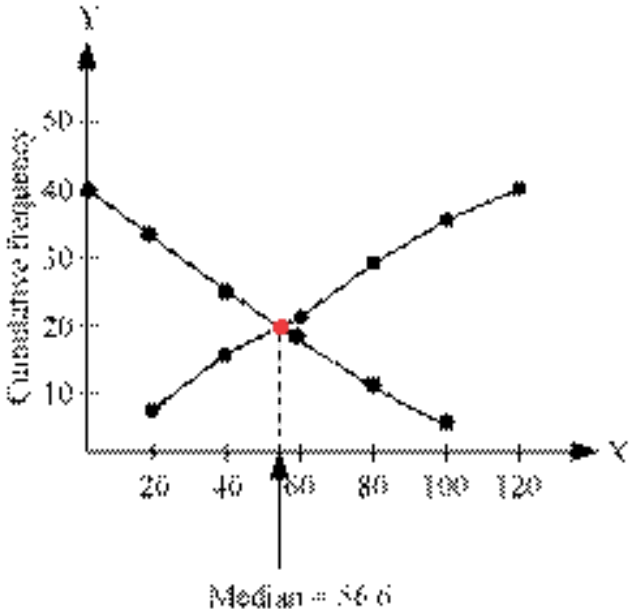
Solution: The cumulative frequency for the given data can be found as:

Class interval	**Lower class limit**	**Frequency**	**Cumulative frequency**
0 – 20	0	7	40
20 – 40	20	8	33
40 – 60	40	6	25
60 – 80	60	8	19
80 – 100	80	6	11
100 – 120	100	5	5

By taking the horizontal axis as the lower class limit and the vertical axis as the corresponding cumulative frequency, we can plot the cumulative frequency for each lower class limit.
Then, the required ogive (of the more-than type) is obtained as:

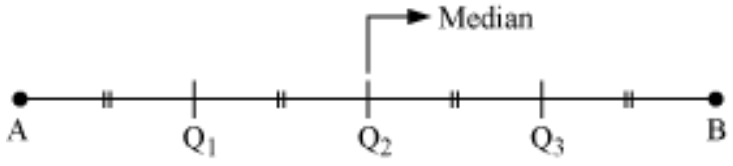


Note:
The x-coordinate of the point of intersection of the "more-than ogive" and "less-than ogive" of a given grouped data gives its median.



- The values of a statistical data which divide the whole set of observations (variate) into four equal parts are known as **quartiles**.

  The points Q_1, Q_2, and Q_3 divide AB into four equal parts.



  Here, Q_2 is median or **middle quartile**. The values Q_1 and Q_3 are respectively called the **lower quartile** and the **upper quartile**.

- In order to find quartiles, firstly, the data is arranged in ascending order of their magnitudes.
- **Lower quartile (Q_1):**

  When the lower half, before the median, is divided into two equal parts, the value of the dividing variate is called lower quartile.

  The formula for calculating the lower quartile is as follows:

$$Q_1 = \begin{cases} \frac{n+1}{4}\text{th observation, if } n \text{ is odd} \\ \frac{n}{4}\text{th observation, if } n \text{ is even} \end{cases}$$

- **Upper quartile (Q_3):**

  When the upper half, after the median, is divided into two equal parts, the value of the dividing variate is called upper quartile.

  The formula for calculating the upper quartile is as follows:

$$Q_3 = \begin{cases} \frac{3(n+1)}{4} \text{th observation, if } n \text{ is odd} \\ \frac{3n}{4} \text{th observation, if } n \text{ is even} \end{cases}$$

- The difference between the upper quartile (Q_3) and the lower quartile (Q_1) is called the **inter-quartile range** and $\frac{\mathbf{Q_3 - Q_1}}{\mathbf{2}}$ is the **semi-quartile range**. Also, inter-quartile range is always positive.